

# Computing the energy of a water molecule using MultiDeterminants: A simple, efficient algorithm

Bryan K. Clark\*

*Princeton Center For Theoretical Science, Princeton University, Princeton, NJ 08544  
Department of Physics, Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544*

Miguel A. Morales†

*Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, California 94550, U.S.A.*

Jeremy McMinis‡

*Department of Physics, University of Illinois at Urbana Champaign, Urbana, IL 61801*

Jeongnim Kim§

*National Center for Supercomputing Applications,  
University of Illinois at Urbana-Champaign, Urbana, IL 61801*

Gustavo E. Scuseria¶

*Department of Chemistry and Department of Physics & Astronomy,  
Rice University, Houston, TX 77005-1892, USA*

Quantum Monte Carlo (QMC) methods such as variational Monte Carlo and fixed node diffusion Monte Carlo depend heavily on the quality of the trial wave function. Although Slater-Jastrow wave functions are the most commonly used variational ansatz in electronic structure, more sophisticated wave-functions are critical to ascertaining new physics. One such wave function is the multiSlater-Jastrow wave function which consists of a Jastrow function multiplied by the sum of Slater determinants. In this paper we describe a method for working with these wavefunctions in QMC codes that is easy to implement, efficient both in computational speed as well as memory, and easily parallelized. The computational cost scales quadratically with particle number making this scaling no worse than the single determinant case and linear with the total number of excitations. Additionally we implement this method and use it to compute the ground state energy of a water molecule.

## I. INTRODUCTION

Being able to accurately calculate the material properties of molecules and solids is important for a variety of fields. There exist a variety of different methodologies to accomplish this including density functional theory [1], quantum chemistry [2] and quantum Monte Carlo [3]. All these methods have different regimes of applicability and different tradeoffs between accuracy and speed. For larger and more complicated systems, quantum Monte Carlo methods are often a good choice due to their polynomial scaling with particle number and high levels of accuracy. For ground state calculations, the most commonly used QMC method is fixed node diffusion Monte Carlo (FNDMC). FNDMC takes as input a trial wave function  $\Psi_T$  and returns the ground state energy of the fixed node wave-function  $\Psi_{FN}$ , the wave-function with the lowest ground state energy with the same nodes as  $\Psi_T$  (i.e.  $\Psi_T(R) = 0 \iff \Psi_{FN}(R) = 0$ ). The accuracy of FNDMC is dependent on the quality of the trial wave function.

Although the most commonly used form of a trial-wavefunction is the Slater-Jastrow form, achieving higher levels of accuracy require the use of more sophisticated ansatz. This is especially true in intrinsically multi-reference problems. Such ansatz are only useful if they can be used with limited computational complexity and memory. A natural extension of the Slater-Jastrow wave-function is the multiSlater-Jastrow form. One advantage of this wavefunction is that any state can be represented in the limit of a large enough number of determinants. Therefore, an answer

---

\*Electronic address: bclark@princeton.edu

†Electronic address: moralesilva2@llnl.gov

‡Electronic address: jmcminis@illinois.edu

§Electronic address: jnkim@illinois.edu

¶Electronic address: guscus@rice.edu

can be systematically improved upon by including more determinants in the ansatz. In this work, we describe a new algorithm that allows FNDMC to use the multiSlater-Jastrow form in an efficient manner. To use a wave-function in FNDMC, there are two important and computationally demanding requirements: the ability to compute ratios of the wavefunction evaluated in two configurations that differ by the location of a single particle and the ability to evaluate gradients and laplacians of the wavefunction. Here we will describe algorithms to accomplish both these tasks. These algorithms will require a minimal amount of additional memory with respect to the single determinant case and will scale (for a single step) with a computational complexity that goes as  $O(n^2 + n_s n + n_e)$  where  $n$  is the total number of particles,  $n_s$  the total number of single excitations and  $n_e$  is the total number of excitations. This scaling is the same with respect to particle number as the single determinant case and scales linearly with the total number of excitations. Other recent work has also proposed an algorithm for computing ratios of multi-determinants [4]. The method described here scales better by a factor of the particle number  $n$  in almost all regimes of interest. In addition, it is significantly simpler requiring less book-keeping and eliminating the need for recursive trees. This simple structure also makes it transparent that the computation involved can be efficiently parallelized. In section II we more concretely introduce the wave function and the necessary operations that must be performed. In Section III we describe an efficient way of performing one of these operations: evaluating wave function ratios. In section IV we focus on computing gradients and laplacians of these terms. Section V discusses other important implementation considerations. In section VI we use our method, implemented in the code QMCPACK [5], on the water molecule. Finally section VII summarizes our results.

## II. BACKGROUND

Fixed Node Diffusion Monte Carlo (FNDMC) take as input a variational wave function. In this paper we discuss the multiSlater-Jastrow which takes the form

$$\Psi = \exp[-J(R)] \sum_k \alpha_k \det M_{\uparrow k} \det M_{\downarrow k} \quad (1)$$

where  $J$  is a Jastrow function and  $\alpha_k$  the weight of the  $k$ 'th determinant. The  $(i, j)$  element of matrix  $M_k$  is equal to

$$M_{\sigma k(i, j)} = \phi_j(r_i) \quad (2)$$

where  $\phi_i$  are 3 dimensional single particle orbitals (s.p.o) selected from the set  $O \equiv \{\phi_1 \dots \phi_{\{n+m\}}\}$  where  $n$  is the total number of electrons and  $m$  is the number of ‘‘virtual orbitals.’’ Although each matrix  $M_{\sigma k}$  may, in principle, contain an arbitrary set of orbitals selected from  $O$ , we consider the typical case where the matrices differ by the action of particle-hole excitations on a reference determinant (i.e. matrices that show up in the multideterminant expansion differ from the reference determinant by at most  $K$  columns.) Replacing one (two, three) orbitals in the reference determinant is respectively called a single (double, triple) excitation. An example where such an expansion naturally arises is the configuration interaction approach in quantum chemistry.

By convention, notate the orbitals in the reference determinant as  $\phi_1 \dots \phi_n$ . The particle-hole excitations are then created by replacing orbitals from the reference determinant with orbitals selected from the  $m$  virtual orbitals. Notate any orbitals in the reference determinant that might be replaced to fill a matrix  $M_{\sigma k}$  for  $k > 0$  as the ground state orbitals. The other orbitals in the reference determinant (typically called frozen or core orbitals) will therefore show up in all determinants in the expansion and can be ignored. We should note that the canonical Slater-Jastrow form simply takes  $\alpha_k = 0$  for all  $k > 0$ .

In quantum Monte Carlo calculations, particles are moved one at a time. We will call the movement of each particle a step. The movement of every particle once is then considered a sweep. There are two basic operations involving the wave-function in FNDMC. They each have to be evaluated once per step (although for the laplacian term, one can also evaluate it once per sweep for every particle  $i$ ). Unless explicitly stated otherwise, the computational complexities discussed in this paper will be the cost per step. The two operations are evaluating the ratio of two wavefunctions that differ by the location of a single coordinate

$$\frac{\Psi_T(r_1 \dots r'_i \dots r_n)}{\Psi_T(r_1 \dots r_i \dots r_n)} \quad (3)$$

and evaluating the quantities

$$\frac{\nabla_i \Psi_T}{\Psi_T}, \frac{\nabla^2 \Psi_T}{\Psi_T} \quad (4)$$

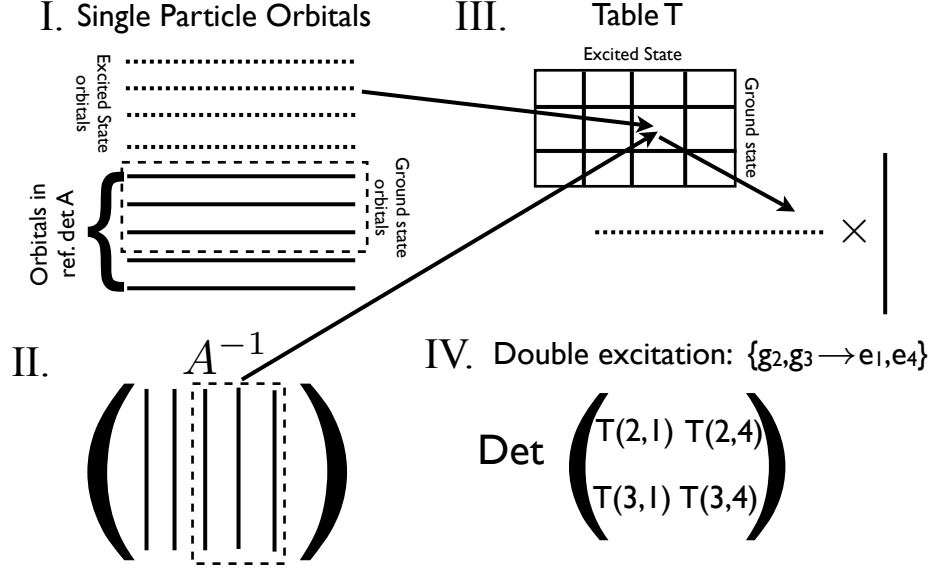


Figure 1:

A graphical representation of the key parts of the MDS algorithm. A table  $T$  (step III) is computed from the dot products of orbitals selected from  $A^{-1}$  (step II - stored throughout the simulation) and the excited state orbitals (step I). Excitations are then read off (step IV) as small determinants of entries from the table  $T$ .

We recall that the wave-function decomposes into the product of two pieces: the Jastrow factor and the multideterminant sum. For computing these quantities, it is sufficient to be able to individually compute the properties for the Jastrows and the Multideterminant expansions. Throughout the rest of this paper, we will focus on evaluating these quantities for the sum of Slater determinants. As the Jastrow factor terms can be dealt with using standard methods, we will not discuss them further.

### III. EVALUATING WAVE FUNCTION RATIOS

In this section we describe a simple, fast procedure for computing ratios of the multideterminant sum (MDS) evaluated on configurations  $R \equiv \{r_1 \dots r_i \dots r_n\}$  and  $R' \equiv \{r_1 \dots r'_i \dots r_n\}$ . The key features of this approach involve only needing to store a single inverse (that of the reference determinant), as well as being able to precompute all the necessary information to generate the excitations at once. Fig. III graphically shows the key steps in this process.

Because a large number of determinant ratios that differ in multiple rows (and occasionally columns) need to be computed, the basic overall approach will involve using the Sherman Morrison-Woodbury formula in evaluating them. We will find that for multiple excitations, though, the values involved in applying these formulas will be needed over and over again and so significant savings can be garnered by intelligently caching intermediate results.

For computing these ratios, the following quantities at each Monte Carlo step (currently located at configuration  $R$ ) must be stored: the inverse of the reference determinants  $M_{\sigma_0}^{-1}(R)$  as well as, for each  $k$ , the determinant  $\det M_{\sigma k}(R)$ . At the beginning of the simulation, the inverse and determinant are calculated from scratch (although even in the initial step, the determinants can be calculated faster using some steps of the procedures described below). Because this only happens once, it does not effect the computational complexity of our procedure.

Given that we store (and update) the determinants so that we always have access to  $\det M_{\sigma k}(R)$  and  $\det M_{\sigma k}(R')$  it should be clear that to evaluate the determinant ratio

$$\frac{\sum_k \alpha_k \det M_{\uparrow k}(R') \det M_{\downarrow k}(R')}{\sum_k \alpha_k \det M_{\uparrow k}(R) \det M_{\downarrow k}(R)} \quad (5)$$

is a straightforward matter of summing and dividing the appropriate quantities each weighted by  $\alpha_k$ . In the rest of this section, then, we focus on the process for updating both  $M_{\sigma_0}^{-1}$  and (for all  $k$ )  $\det M_{\sigma k}$  as we move from  $R \rightarrow R'$ .

In moving from  $R \rightarrow R'$  the reference determinants  $M_{\sigma_0}^{-1}(R')$  needs to be evaluated (recalling we already start with  $M_{\sigma_0}^{-1}(R)$ ). This is exactly the same requirement for the typical Slater-Jastrow wave-functions and will be accomplished

in the canonical manner using the Sherman-Morrison formula. Because this type of update will be used throughout this paper in a variety of different contexts, we will outline the process here for completeness. First note that since  $R$  and  $R'$  differ by a single particle,  $r_i$ , this means that  $M_{\sigma 0}(R)$  and  $M_{\sigma 0}(R')$  differ by the change of a single row. Using the Sherman-Morrison formula one can update the inverse of a matrix  $M_{\sigma 0}(R')$  which differs from  $M_{\sigma 0}(R)$  via a single row by computing

$$M_{\sigma 0}^{-1}(R') = M_{\sigma 0}^{-1}(R) - \frac{[M_{\sigma 0}^{-1}(R)]u e_k^T [M_{\sigma 0}^{-1}(R)]}{1 + e_k^T M_{\sigma 0}^{-1}(R)u} \quad (6)$$

where  $u_j = \phi_j(r'_i) - \phi_j(r_i)$  and  $e_k$  is the Cartesian basis vector with a 1 at position  $k$ . In the process of computing this, it should be noted that the determinant ratio

$$\frac{\det M_{\sigma 0}(R')}{\det M_{\sigma 0}(R)} = 1 + e_k^T M_{\sigma 0}^{-1}(R)u \quad (7)$$

has already been computed from the denominator of our update. Generically evaluating this latter quantity involves computing a single dot product and costs  $O(n)$  time. Since  $\det M_{\sigma 0}(R)$  is already stored, one can compute  $\det M_{\sigma 0}(R')$  from this ratio. This total operation of updating  $M_{\sigma 0}^{-1}$  and  $\det M_{\sigma 0}$  is an  $O(n^2)$  operation.

In addition to updating the reference determinant the new terms  $\det M_{\sigma k}(R')$ ,  $k > 0$  need to be evaluated for the new configuration  $R'$ . These terms will be generated by taking the product of  $\det M_{\sigma 0}(R')$  (computed from the denominator of the inverse updates) and the ratio  $\det M_{\sigma k}(R')/\det M_{\sigma 0}(R')$ . Computing these latter ratios will be the focus of the rest of this section.

It should be noted that a naive approach to computing each ratio  $\det M_{\sigma k}(R')/\det M_{\sigma 0}(R')$  involves storing  $M_{\sigma k}^{-1}(R)$  for each  $k$  and updating it as we have done for the reference determinant. If we let  $n_e$  be the total number of determinants this gives us  $O(n_e n^2)$  time per step making the use of a MDS cost  $n_e$  times a single determinant. This is a significant cost and consequently instead of computing the ratios  $\det M_{\sigma k}(R')/\det M_{\sigma 0}(R')$  in this manner, we will do something more sophisticated.

Recall that in computing the series of ratios  $\det M_{\sigma k}(R')/\det M_{\sigma 0}(R')$ , we start with the inverse matrix  $M_{\sigma 0}^{-1}(R')$  where  $M_{\sigma k}$  and  $M_{\sigma 0}$  differ by  $s$  columns (and hence is an excitation of  $s$  particle-hole pairs). The general approach for computing these quantities will be as follows: we fill a table of elements and then compute the desired ratios by evaluating determinants of small  $(s \times s)$  matrices whose elements are taken from this table. Each element  $(i, j)$  in the table will be generated via the dot product of the single particle orbitals  $\phi_i(R')$  with a column of the matrix inverse  $[M_{\sigma 0}^{-1}(R')]_j$ .

In more detail, let  $\{g_1 \dots g_k\}$  be the set of all orbitals in the reference matrix  $M_0(R')$  that must be replaced by orbitals  $\{e_1 \dots e_m\}$  to produce all the excitation matrices  $M_1(R') \dots M_{n_e}(R')$  in the MDS. Additionally let  $\{g_1^{-1} \dots g_k^{-1}\}$  be the corresponding columns in the inverse matrix  $M_0^{-1}(R')$ . An example double excitation then might replace  $\{g_1, g_5\} \rightarrow \{e_2, e_7\}$ .

To compute the needed ratios, first, generate a table  $T$  of size  $k \times m$  where the  $(i, j)$  element of the table is  $g_i^{-1} \cdot e_j$ . Note that each dot product costs  $O(n)$  and one must do this for  $km$  different elements in our table giving us a cost of  $O(kmn)$ . For each single excitation of the form  $e_i \rightarrow g_j$  read off the  $T(i, j)$  element of the table. Notice that if the MDS spanned all possible pairs of single particle excitations from  $\{g\} \rightarrow \{e\}$  (i.e. there are  $n_s = km$  single excitations), then every element computed in the table  $T$  is utilized. This is a typical occurrence and means all the computation so far was necessary even if only single particle excitations are calculated.

The work required for excitations above the single excitation (double, triple, etc.) is constant per excitation. For each double excitation of the form  $(e_i, e_j) \rightarrow (g_k, g_l)$ , compute the determinant of the  $2 \times 2$  matrix

$$\det \begin{pmatrix} T(i, k) & T(i, l) \\ T(j, k) & T(j, l) \end{pmatrix} \quad (8)$$

More generically, to compute an excitation replacing  $r$  excitations with  $r$  ground state orbitals, compute the determinant of the  $r \times r$  matrix where the rows are labelled by the excitation orbitals, the columns by the ground state orbitals and the matrix elements selected from their corresponding orbital pairs in  $T$ .

Notice, that computing a  $r'$ th level excitation now simply involves reading off  $r^2$  matrix elements and computing the determinant of a  $r \times r$  matrix ( $O(r^3)$  time). Consequently, this step takes

$$O\left(\sum i^3 n_i\right) \quad (9)$$

---

**Algorithm 1** Fast MultiDeterminant when moving from  $R \rightarrow R'$ 


---

1. Update the reference determinants inverse  $M_{\sigma 0}^{-1}(R')$  using the Sherman-Morisson formula. Cost:  $O(n^2)$
  2. Update the single particle orbitals of the excited states  $\{e_1 \dots e_m\}$ . Cost:  $O(mn)$
  3. For all ground state orbitals  $\{g_1, g_2 \dots g_k\}$  whose corresponding row in  $M_{\sigma 0}^{-1}(R')$  is  $\{g_1^{-1}, g_2^{-1} \dots g_k^{-1}\}$  and all excited states  $\{e_1 \dots e_m\}$  compute a  $k \times m$  size table  $T$  whose  $(i, j)$  element is  $g_i^{-1} \cdot e_j$ . This can be done by  $km$  dot products or a single matrix multiplication. Cost:  $O(kmn) = O(n_s n)$
  4. To evaluate the ratio of  $\Psi(\{g\} \rightarrow \{e\})/\Psi(\{g\})$ , compute the determinant of a matrix whose elements are selected from  $T$  corresponding to the rows and columns that make up all pairs in  $\{g\}$  and  $\{e\}$  Cost:  $O(n_e)$
- 

time where  $n_i$  is the number of  $i$ -th level excitations. Since  $r$  is almost always bounded by some small value less than 8 (rarely are octuple excitations considered), this becomes an operation proportional to the number of excitations  $O(n_e)$ .

In the situation where we span all possible single excitations, this gives us a total cost per step (including updating the reference determinant) of

$$O(n^2) + O(kmn) + O(n_e) = O(n^2) + O(n_s n) + O(n_e) \quad (10)$$

where  $n_s$  is the number of single excitations and  $n_e$  the total number of determinants. Since the cost per step for evaluating a Slater-Jastrow wave function scales as  $O(n^2)$  this gives us an additional cost of  $O(n_s n) + O(n_e)$  to use multideterminant wave functions. Algorithm 1 summarizes the cost of these different steps.

We can also compute the memory cost for using this approach. We need to store a  $n \times n$  matrix (as in the single determinant case) to store the inverse  $M_{\sigma 0}^{-1}$ . We also need to store the single particle orbitals evaluated on all the particles. These cost  $n^2 + (m + n)n$  memory and need to be stored throughout the simulation. At each step, we also need to produce the table  $T$  of size  $km = n_s$  and store the values of the determinant ratios ( $\sim n_e$ ). This gives a complete cost in memory of  $(n^2 + (m + n)n + n_s + n_e)$ . We note that the primary cost in terms of memory is dominated by storing the single particle orbitals. As is often the case, there is some tradeoff between memory and computational complexity and the total memory could be decreased at the cost of some recomputation.

#### IV. GRADIENTS AND LAPLACIANS

In FNDMC, in addition to computing ratios, the terms

$$\frac{\nabla_i \Psi}{\Psi} \quad (11)$$

and

$$\frac{\nabla_i^2 \Psi}{\Psi} \quad (12)$$

also need to be evaluated. The gradients must be evaluated at each step while the laplacian is required for evaluating the kinetic energy and can be evaluated at each step or after one sweep (i.e. a step of each particle). In the following section we will write  $\nabla_i^\alpha$  to notate one of  $(\nabla_i^x, \nabla_i^y, \nabla_i^z, \nabla_i^2)$  and will use the term gradients as shorthand for gradients and laplacian. We assume throughout the rest of this section that we are only working with one specific particle  $i$ .

As calculating the Jastrow's can be done in the standard way, we again focus on the gradients of the multi-determinant term. Such terms look like

$$\frac{\nabla_i^\alpha \sum_k \alpha_k \det M_{\uparrow k} \det M_{\downarrow k}}{\sum_k \alpha_k \det M_{\uparrow k} \det M_{\downarrow k}} \quad (13)$$

The denominator of eqn. 13 shows up in evaluating the ratios, and, as such, the machinery necessary to compute this has already been developed. To compute the numerator, we evaluate the quantities  $\nabla_i^\alpha \det M_{\sigma k}$  (from which the numerator is then easily calculable) by evaluating

$$\frac{\nabla_i^\alpha \det M_{\sigma k}}{\det M_{\sigma 0}} \quad (14)$$

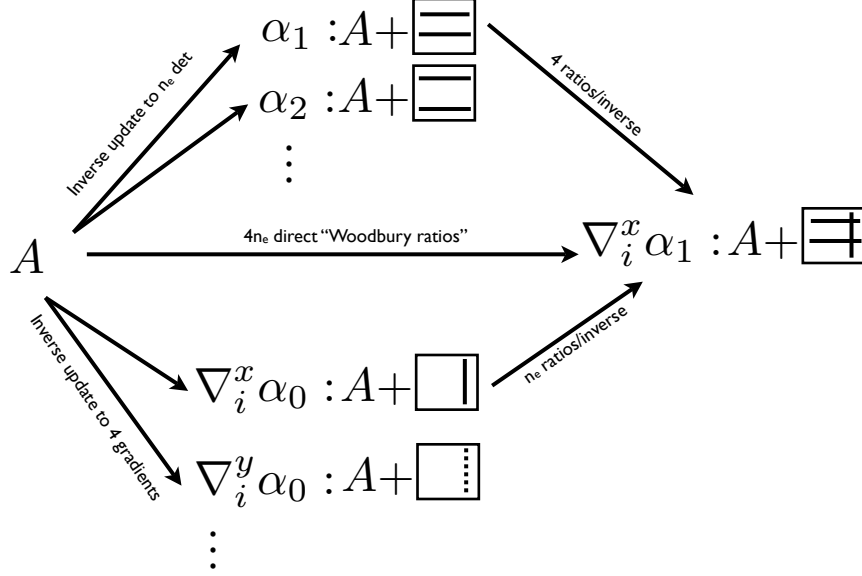


Figure 2:

A graphical representation of the different approaches to computing the derivative of the matrix that differs by multiple rows and 1 column. The most efficient path (in most regimes) is to follow the lower one.

For a given particle  $i$  and each of the four  $\nabla_i^\alpha$  terms, eqn. 14 can be written as the determinant ratio between two matrices:  $M_{\sigma 0}$  and a matrix generated by replacing  $s$ -columns (for a  $s$  particle excitation) and 1 row (corresponding to particle  $i$ ) from  $M_{\sigma k}$ . Because the two matrices differ in *both* rows and columns, the typical approaches for computing ratios can not be straightforwardly applied. Instead there are three possible approaches for dealing with this. Fig. IV A outlines these three options.

#### A. Updating the gradients first

Recall that we start our evaluation with the reference matrix inverse  $M_{\sigma 0}^{-1}$  and need to compute the ratio of two determinants that differ in both rows and columns. We will compute the ratio of these two determinants in two steps by working with an intermediate matrix for each of the four  $\alpha$ . This intermediate matrix which we notate as  $M_{\sigma 0}^\alpha$  corresponds to replacing the elements of the  $i$ 'th row of  $M_{\sigma 0}$  whose  $j$ 'th element currently is  $\phi_j(r_i)$  with the values (for their  $j$ 'th element) corresponding to  $\nabla_i^\alpha \phi_j(r_i)$ . It should be noted that

$$\frac{\det M_{\sigma 0}^\alpha}{\det M_{\sigma 0}} = \frac{\nabla_i^\alpha \det M_{\sigma 0}}{\det M_{\sigma 0}} \quad (15)$$

For each of these 4 matrices, we explicitly compute its inverse  $[M_{\sigma 0}^\alpha]^{-1}$  using the Sherman Morrison formula. This costs  $O(n^2)$  work. Recall, as part of this computation, this also gives us the ratio  $\det M_{\sigma 0}^\alpha / \det M_{\sigma 0}$ . Notice that, computing  $\det M_{\sigma k}^\alpha / \det M_{\sigma 0}^\alpha$  for a given  $k$  would then be sufficient to compute eqn. 14 for that  $k$ . These two matrices, though, differ from each other by the same series of particle-hole excitations that we worked with in the ratios. In order to compute  $\det M_{\sigma k}^\alpha / \det M_{\sigma 0}^\alpha$  for all  $k$  we will do exactly the same thing as we did for the determinant ratios! The only difference in this case is when building the table  $T$ , each element of the sets  $\{e\}$  have their  $i$ 'th value replaced by the gradient of that value and the elements in the set  $\{g\}^{-1}$  now differ because the reference inverse is different. It is clear then that the cost of this operation is four times the cost of evaluating a determinant ratio. Therefore, computing all the gradients and laplacians for a single particle costs  $4(n^2 + nm_s + n_e)$ . This means that computing the gradients and laplacians twice per step (the amount required for FNDMC) is only a constant factor more costly then computing the determinant ratios themselves. Although this is worse then the single determinant case (where the ratios can all be computed without updating any inverses) the total computational complexity is still well controlled.

## B. Updating each determinant first

As an alternative approach, instead of producing an updated inverse for each of the four  $\nabla_i^\alpha$  we instead could have produced an updated inverse for each multi-determinant matrix and then do row-ratios with respect to the four  $\nabla_i^\alpha$  we need. For the computation of the gradient of a single particle this is always an inferior approach. If the gradient is being computed for all particles simultaneously, then in this alternative approach, the updated inverse needs only be computed once and so its cost can be amortized over all particles. (No such savings for doing all particles simultaneously can be garnered in the table method). Working out the cost for computing all the terms (i.e. one sweep) gives a cost of  $O(n^2 n_e)$ . This is computationally superior in the situation where  $n^2 n_e < n^3$ . We expect such situations to be rare but in these cases and when all particles need to be calculated simultaneously, this latter approach scales asymptotically better.

## C. Direct Application of the Generalized Matrix Determinant Lemma

The generalized matrix determinant lemma states

$$\det(A + UV^T) / \det A = \det(I + V^T A^{-1} U) \quad (16)$$

There exists choices of  $U$  and  $V$  (taking matrices of size  $(s+1) \times n$ ) that allow  $A + UV^T$  and  $A$  to differ by  $s$  columns and 1 row. Then, for each  $i$  and each  $\alpha$ , one could directly apply this lemma. An upper bound on this cost per  $(i, \alpha)$  is  $(4n^2 + 16n)$  (the cost can be brought down somewhat because some of the matrix vector multiplications involved are against cartesian basis vectors and hence can be computed quickly). Performing this computation for the gradients and laplacians of all the excitations will then take  $16n^2 n_e$ . This is prohibitively expensive and not the correct approach when computing the gradients or laplacians for all the particles. Nonetheless, it might be a useful method to have in the rare case where a single term needs to be computed. Additionally, this has the added benefit of not having numerical instability problems that can potentially crop up in other approaches.

## V. ADDITIONAL CONSIDERATIONS

In this section, we present a couple of additional considerations that should be considered when implementing the Table method. Specifically these involve efficiency considerations, numerical instabilities, and parallelization.

To begin with, the dominating effort in computing determinant ratios involves the computation of dot products or rows/columns of matrices. The performance of the entire algorithm will depend heavily on optimizing the application of these dot products. Specific care should be taken to ensure that the dot products are taken so the fastest index is being looped over. Additionally, in a number of cases, these dot products can be chunked together into a single large matrix multiplication step (as opposed to a series of dot products). For example, in computing the elements of the table, instead of computing all pairs of dot products between the sets of  $\{g\}^{-1}$  and  $\{e\}$  one can instead compute all these dot products simultaneously as a matrix multiplication between two matrices whose rows/columns are respectively the vectors of  $\{g\}^{-1}$  and  $\{e\}$ . Not only will this result in significantly better cache performance, and an ability to take advantage of highly-tuned matrix multiplication routines, but formally [19] can also be asymptotically faster.

Concerning numerical instabilities, there are two places in this algorithm where such instabilities can crop up. (Although these are both theoretically possible, we never find them to be a problem in practice and so have not yet been forced to implement the suggested methods to avoid them). To begin with, the inverse of the reference determinant  $M_0$  could become (close to) singular. When evaluating wave functions of the Slater-Jastrow type this is never a problem because if  $M_0$  becomes (close to) singular, the probability of moving to that configuration is essentially 0 and so the move is rejected. This can not be ensured in the case of multiSlater-Jastrow. It could be the value of the reference matrix goes to 0 but since the other determinants have a non-zero value, the move itself would be accepted. This problem can be dealt with by temporarily switching to another of the determinants in our expansion as the reference determinant. This increases the cost slightly as the size of all the excitations may potentially increase by the number of particle-hole excitations in the new reference determinant, but this new reference determinant need only be used for small portions of the simulation.

The second area where numerical instabilities might appear is in the calculation of the gradients. Again it is possible in the inverse update step, that the matrices end up being nearly singular. Of course, again one could potentially find an appropriate new reference determinant. Because an update step is not strictly required in this case, though, another option exists. Instead of updating an inverse at all, for each multi-determinant, we can evaluate the ratio

with respect to both rows and columns simultaneously using the generalized matrix-determinant lemma (as described in section IV-C). This can be done in time  $O(n^2) + O(nn_e)$  (to achieve even this scaling, one must be careful not to perform identical matrix multiplications twice). We note that this is potentially significantly more expensive than the table method ( $nn_e$  instead of  $nn_s$ ), but needs only to be done upon seeing numerical instabilities.

Finally, it should be pointed out that the algorithms described in this paper can be easily parallelized especially over many cores on a single machine. The table of elements can be easily broken up into chunks each of which can be computed by separate cores. Moreover, once this table has been produced, reading and computing the tiny determinants can be done in parallel. Consequently, using this approach we believe that the evaluation of these multideterminant terms can scale particularly well.

## VI. WATER MOLECULE

In this section, we benchmark our approach on the all electron water molecule at the equilibrium geometry  $r_{OH} = 0.9572 \text{ \AA}$ ,  $\theta_{HOH} = 104.52^\circ$ . There are accurate results for this molecule and it has received considerable attention from the QMC community in the past [6–9]. Since we do not employ orbital optimization [10] in our QMC calculations, we rely on high level quantum chemistry methods to produce a good set of molecular orbitals for our multideterminant expansion. All quantum chemistry calculations were performed with the GAMESS code [11], using the Roos augmented triple zeta ANO Gaussian basis set [12–14]. To obtain the molecular orbitals, we start with a Complete Active Space Self-Consistent Field (CASSCF) calculation including all 10 electrons in 8 orbitals. The resulting orbitals are used in a Multi Configuration Self-Consistent Field (MCSCF) calculation including all single and double excitations from the CAS involving the next 44 orbitals; this is known as second-order CI (SOCi) in the chemistry community. We used the natural orbitals of the converged MCSCF calculation and choose the configurations in our expansion by applying a cutoff to the SOCi wave function. We found this approach to balance the complexity of the quantum chemistry calculation with the quality of the resulting set of orbitals and configurations. Instead of using determinants directly in our calculations, we use configuration state functions (CSF) which are spin and space adapted linear combination of Slater determinants. This eliminates redundant variational parameters from the wave function and facilitates the optimization process. To this multideterminant expansion we add a Jastrow factor that contains 1,2, and 3 body terms and optimize all the variational parameters simultaneously (including the nonlinear Jastrow coefficients and the linear CSF coefficients) by energy minimization based on a variant of the linear method of C. Umrigar, et. al. [10].

Figure 3 shows the speedup (ratio of the time required between the base and Table methods) obtained with the Table method in the evaluation of the ratios and gradients of the list of determinants involved in the construction of the wavefunction during a VMC move. The blue line does not include the time spent performing the inverse updates in the canonical algorithm (nor the inverse update of the reference determinant in the Table algorithm). For the comparisons including inverse updates two time steps are shown, giving approximately a 50% and 100% acceptance rate in the VMC calculation. These cases represent the two typical limits encountered in QMC. Although the exact speedup depends on the details of the particular problem, with a few thousand determinants we get speedups on the order of 25. For fixed  $n$  (and fixed or small  $n_s$ ), both the canonical and Table algorithms scale as  $O(n_e)$  and therefore for large  $n_e$  we expect speedup to be a constant factor as we see in Figure 2. Because the prefactor before this constant goes as  $n^2$  for the canonical algorithm and constant for the Table method, we expect the speedup to grow for larger systems. Oscillations seen before reaching this asymptotic limit, is caused by implementation details, e.g. cache usage, locality of references in memory, etc. These timings are not meant to be a reflection of the operation counts involved in the algorithm, but merely an example of the increased efficiency of our implementation in QMCPACK.

Figure 4 shows our VMC and DMC total energies of the water molecule as a function of the sum of the squares of the initial CSF coefficients (those produced by the MCSCF calculation). The latter is related to the fraction of important configurations (according to the MCSCF calculation) included in the expansion, as we approach unity we include all the determinants in the SOCi wavefunction. Our results not only show the ability of this wavefunction to recover large amounts of correlation, but also the systematic improvement of the energy as more configurations are included. Notice that a stable and reliable optimization method is needed in order to recover the maximum amount of correlation, the details of our optimization algorithm will be discussed in a future publication [15]. Table 1 shows a comparison of our calculations with a selection of previous results. Our best wavefunction (including 3461 CSF) recovers 99.7% of the correlation energy, with an error of 1.2 mHa. It should be noted that the energy computed in DMC with the determinants selected from the SOCi wavefunction is significantly better than the energy that comes out of the SOCi calculation directly,  $E_{SOCi} = -76.3619395110$ .



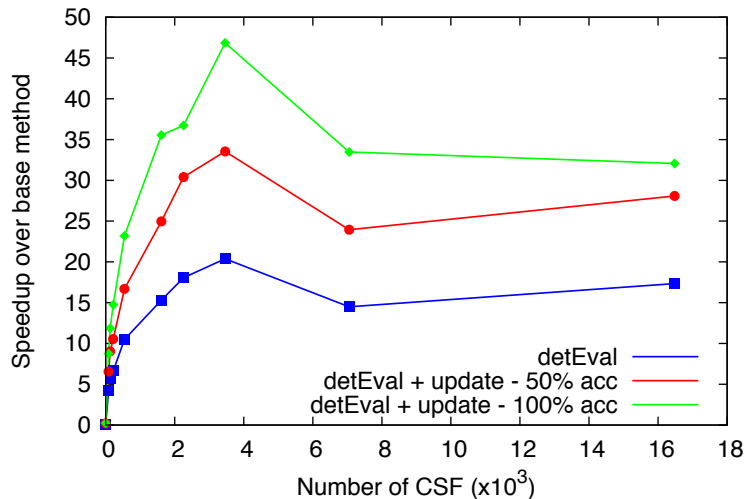


Figure 3: Speedup obtained with the Table method in the evaluation of the ratios and gradients of the list of determinants involved in the construction of the wavefunction during a VMC move. The blue line only includes the calculation of the ratios and gradients, while the other lines also include the inverse update. The red and green lines use time steps that produce acceptance rates of 50% and 100% respectively.

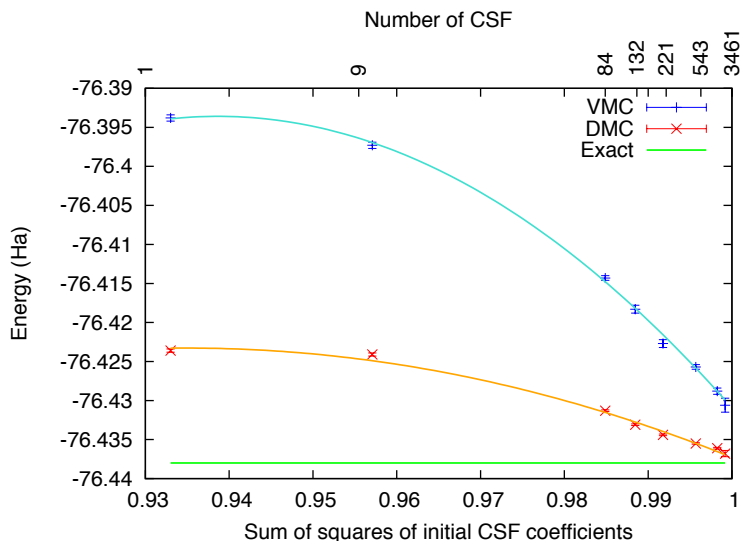


Figure 4: Total energy of the water molecule as a function of the sum of squares of the initial CSF coefficients.

## VII. CONCLUSIONS

Although multideterminant expansions were introduced in QMC more than a decade ago, their use in molecular problems has been uncommon and the number of configurations typically included in calculations is rather small. In this work we have described an algorithm that allows for their efficient evaluation. Specifically we focus on the steps of computing wave function ratios and evaluating gradients and laplacians of these wave functions. Beyond the typical inverse update of a reference matrix that is required for just the Slater-Jastrow ansatz, the algorithm for evaluating ratios involves two key steps: computing a table (whose elements are often all computable by a single matrix multiplication) and then reading off ratios from elements (or determinants of a few elements) of this table. Gradients and laplacians can be computed in the same way as ratios where the reference matrix being used is the matrix where a row is replaced by its gradient (respectively laplacian). This procedure allows the use of thousands of determinants for a cost that is only a few times the cost of a single determinant. We test this empirically on a water dimer getting good energies at a reasonable computational cost. This algorithm will open up the possibility of

Method	Total Energy (Ha)	Ref.
VMC-SingleDet	-76.3938(4)	This work.
VMC-MSDJ	-76.4306(9)	This work.
VMC-SJB	-76.4034(2)	[7]
DMC-SingleDet	-76.4236(2)	This work.
DMC-MSDJ	-76.4368(4)	This work.
DMC-B3LYP	-76.4230(1)	[7]
DMC-SJB	-76.42830(5)	[7]
DMC-AGP	-76.4175(4)	[6]
DMC-PNO-CI	-76.429(1)	[9]
CCSD(T)-R12	-76.4373	[16]
CEEIS	-76.4390(4)	[17]
Exact	-76.438	[18]

Table I: Comparison of results on the water molecule. SingleDet refers to our calculations employing only a single determinant and MSDJ refers to our multiSlater-Jastrow wavefunction.

retrieving a significant percent of the correlation energy in a variety of strongly correlated systems.

### Acknowledgments

We would like to thank Ken Esler for useful conversations and Cyrus Umrigar for carefully reading and commenting on the manuscript. The work at Rice University was supported by the Department of Energy (DE-FG02-04ER15523) and the Welch Foundation (C-0036). This work was performed in part under the auspices of the US DOE by LLNL under Contract DE-AC52-07NA27344. The work at UI was supported by the National Science Foundation under No. 0904572 and EFRC - Center for Defect Physics sponsored by the US DOE, Office of Basic Energy Sciences.

- 
- [1] R. Martin, *Electronic structure: basic theory and practical methods* (Cambridge Univ Pr, 2004), ISBN 0521782856.
  - [2] A. Szabo and N. Ostlund (1996).
  - [3] W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, *Mod. Phys* **73**, 33 (2001).
  - [4] P. Nukala and P. Kent, *The Journal of chemical physics* **130**, 204105 (2009).
  - [5] <http://qmcpack.cmscc.org> (2011).
  - [6] M. Casula, C. Attaccalite, and S. Sorella, *The Journal of chemical physics* **121**, 7110 (2004).
  - [7] I. G. Gurtubay and N. R. J., *The Journal of chemical physics* **127**, 124306 (2007).
  - [8] N. A. Benedek, I. K. Snook, M. D. Towler, and N. R. J., *The Journal of chemical physics* **125**, 104302 (2006).
  - [9] A. Luchow and R. F. Fink, *The Journal of chemical physics* **113**, 8457 (2000).
  - [10] J. Toulouse and C. J. Umrigar, *The Journal of chemical physics* **128**, 174101 (2008).
  - [11] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, et al., *J. Comput. Chem.* **14**, 1347 (1993).
  - [12] P. Widmark, P. Malmqvist, and B. Roos, *Theor. Chim. Acta* **77**, 291 (1990).
  - [13] K. Schuchardt, B. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li, and T. Windus, *J. Chem. Inf. Model.* **47**, 1045 (2007).
  - [14] D. J. Feller, *J. Comp. Chem.* **17**, 1571 (1996).
  - [15] Wavefunction optimization, In preparation. (2011).
  - [16] H. Muller and W. Kutzelnigg, *Molecular Physics* **92**, 535 (1997).
  - [17] L. Bytautas and K. Ruedenberg, *The Journal of chemical physics* **124**, 174304 (2006).
  - [18] D. Feller, C. M. Boyle, and E. R. Davidson, *The Journal of chemical physics* **86**, 3424 (1986).
  - [19] Matrix multiplication is known to scale theoretically better than  $O(n^3)$  (i.e. at least  $O(n^{2.376})$ ) but these asymptotically faster algorithms are rarely useful in practice.